

## Tilburg University

### Mechanistic and non-mechanistic varieties of dynamical models in cognitive science

Gervais, R.R.

*Published in:*  
Synthese

*DOI:*  
[10.1007/s11229-014-0548-5](https://doi.org/10.1007/s11229-014-0548-5)

*Publication date:*  
2015

*Document Version*  
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Gervais, R. R. (2015). Mechanistic and non-mechanistic varieties of dynamical models in cognitive science: Explanatory power, understanding, and the 'mere description' worry. *Synthese*, 192(1), 43-66.  
<https://doi.org/10.1007/s11229-014-0548-5>

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# *Mechanistic and non-mechanistic varieties of dynamical models in cognitive science: explanatory power, understanding, and the 'mere description' worry*

**Raoul Gervais**

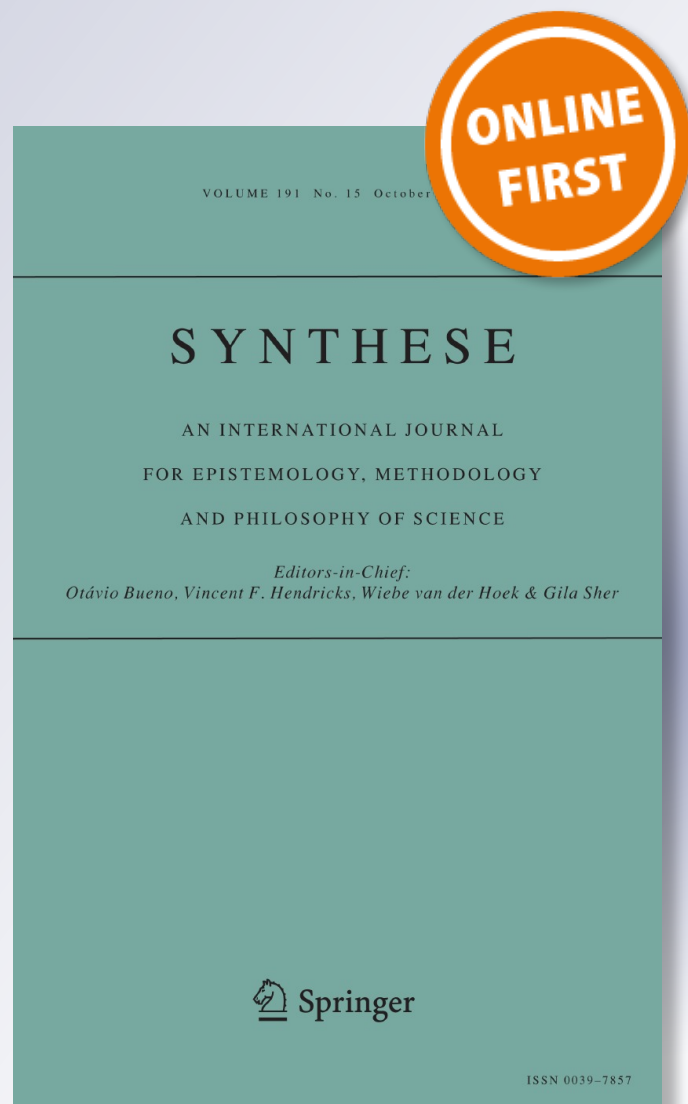
## **Synthese**

An International Journal for  
Epistemology, Methodology and  
Philosophy of Science

ISSN 0039-7857

Synthese

DOI 10.1007/s11229-014-0548-5



**Your article is protected by copyright and all rights are held exclusively by Springer Science +Business Media Dordrecht. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at [link.springer.com](http://link.springer.com)".**

# Mechanistic and non-mechanistic varieties of dynamical models in cognitive science: explanatory power, understanding, and the ‘mere description’ worry

Raoul Gervais

Received: 29 November 2013 / Accepted: 26 August 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** In the literature on dynamical models in cognitive science, two issues have recently caused controversy. First, what is the relation between dynamical and mechanistic models? I will argue that dynamical models can be upgraded to be mechanistic as well, and that there are mechanistic and non-mechanistic dynamical models. Second, there is the issue of explanatory power. Since it is uncontested the mechanistic models can explain, I will focus on the non-mechanistic variety of dynamical models. It is often claimed by proponents of mechanistic explanations that such models do not really explain cognitive phenomena (the ‘mere description’ worry). I will argue against this view. Although I agree that the three arguments usually offered to vindicate the explanatory power of non-mechanistic dynamical models (predictive power, counterfactual support, and unification) are not enough, I consider a fourth argument, namely that such models provide understanding. The Voss strong anticipation model is used to illustrate this.

**Keywords** Dynamical cognitive science · Dynamical models · Mechanism · Explanation · Understanding · Strong anticipation

## 1 Introduction

There is some controversy over the question whether dynamical models in cognitive science are genuinely explanatory, or whether they merely (re)describe cognitive phenomena (Eliasmith 1996; van Gelder 1998; Dietrich and Markman 2001; van

---

R. Gervais (✉)  
Centre for Logic and Philosophy of Science, Ghent university,  
Blandijnberg 2, 9000 Ghent, Belgium  
e-mail: raoul.gervais@ugent.be

Leeuwen 2005; Craver 2007; Walmsley 2008).<sup>1</sup> Closely tied up with this controversy are differences about the explanatory *nature* of such models. In particular, the ‘mere description’ worry arises from the fact that dynamical models are often held to derive their explanatory power from their ability to predict future states of the target cognitive system by inferring them from certain regularities, together with antecedent conditions (van Gelder 1998, p. 625; Chemero and Silberstein 2008; Walmsley 2008)—a view that has been labeled ‘predictivism’ (Kaplan and Craver 2011). However, a well-documented problem with such predictivist accounts of explanation is that predictive success can also be afforded by merely *phenomenal* regularities—that is, by regularities that merely describe, rather than explain. Indeed, predictivism inherits this problem from Hempel’s covering-law (henceforth CL) account of explanation.<sup>2</sup>

The mere description worry is often put forward by the proponents of mechanistic explanations (Bechtel 1998; Craver 2006; Kaplan and Craver 2011; Kaplan and Bechtel 2011). This is not to say that these authors actually believe that dynamical models are not explanatory. Rather, they think that the objection shows that the only way in which dynamical explanations can explain a phenomenon is by “...embracing commitments about the causal mechanisms that produce, underlie, or maintain it” (Kaplan and Craver 2011, p. 603). In a nutshell, only insofar as dynamical models are mechanistic, can they be said to explain. Maybe some dynamical models really do fail to embrace mechanistic commitments, but then such models are not explanatory.<sup>3</sup> Of course, before settling these issues one must first get clear on the relation between dynamical and mechanistic models.

In this article, I will argue for two claims. First, I will argue that dynamical and mechanistic models overlap to a certain extent; more precisely, that while dynamical models are predictivist, they can be augmented or upgraded in a way that makes them mechanistic as well. Second, I will argue that both varieties can be genuinely explanatory. Since it is uncontested the mechanistic models can explain, I will focus on the non-mechanistic variety. Although I agree that the three arguments usually offered to vindicate the explanatory power of non-mechanistic dynamical models, namely predictive power, counterfactual support, and unification, are not enough, I propose a fourth argument, namely that such models provide understanding. In spelling out this notion, I will argue that providing understanding is a necessary condition on explanation. Although ultimately, one may resist the claim that understanding is sufficient for explanation, at the very least the argument shows that non-mechanistic dynamical models are more than mere descriptions.

<sup>1</sup> Of course, the question here is whether dynamical models are *necessarily* non-explanatory, not whether there are examples of dynamical models that happen to be non-explanatory—even those who answer the former question negatively can admit that there might be bad *instances* of dynamical models.

<sup>2</sup> In the words of Zednik: “As long as dynamical explanation is viewed as a form of covering-law explanation [...] the mere description worry looms” (2011, p. 246). Although, as we will see, dynamical models do have CL-type properties, ultimately, they fall far short of Hempel and Oppenheim’s (1948, p. 137) strict requirements. In particular, the generalities employed do not meet the traditional requirements for lawhood. Therefore, I will stick to the term ‘predictivism’.

<sup>3</sup> Although it is not denied that such non-explanatory dynamical models may have other virtues (Kaplan and Bechtel 2011, p. 443).

Here is an overview of the paper. Section 2 consists of a detailed comparison between dynamical and mechanistic models. I will argue that although dynamical models have the predictivist property of inferring their explananda from regularities, there is nothing against supplementing them with commitments about the underlying mechanism. Specifically, the position I will argue for is non-exclusivity: to a certain extent, dynamical and mechanistic models overlap. Having established this, in Sect. 3 I will argue that for the non-mechanistic variety of dynamical models, the mere descriptions worry cannot be overcome by appealing to the traditional arguments proposed by the predictivists—an additional argument is needed. In Sect. 4, I will introduce the notion of scientific understanding, focusing on the contextual theory of understanding developed by De Regt and Dieks (2005). Finally, pace Kaplan and Bechtel (2011), in Sect. 5 I will show how non-mechanistic dynamical models can be explanatory in that they provide the sort of understanding described in Sect. 4, by considering an example of a dynamical model of which both sides agree that it is non-mechanistic, namely Voss' strong anticipation model (2000).

## 2 Some dynamical models are mechanistic, while others are non-mechanistic

### 2.1 Dynamical and mechanistic models: exclusivity versus non-exclusivity

Leaving aside the issue of explanatory power for a moment, how do dynamical and mechanistic models relate? Do these model-types overlap to a certain extent, or are they mutually exclusive? Several positions are possible. Here is a non-exhaustive list, restricted to the domain of cognitive science:<sup>4</sup>

- *E1* All dynamical models are predictivist, and hence non-mechanistic.
- *E2* All dynamical models are mechanistic, and hence not predictivist.
- *E3* Some dynamical models are predictivist, while others are mechanistic.
- *NE1* All dynamical models are predictivist, but some of them are mechanistic as well (by being 'upgraded' to mechanistic models).
- *NE2* All dynamical models are mechanistic, but some of them are also predictivist.
- *NE3* All dynamical models are predictivist as well as mechanistic.

The first three positions are exclusive, in that they assume that a model cannot be both dynamical and mechanistic, while the final three are non-exclusive, in that they assume that a model can be both. Over the course of this section, I will argue that *NE1* is correct; in other words, that there is a partial overlap between dynamical and mechanistic models. As we will see, the predictivist dynamical models can be augmented or interpreted so as to become mechanistic—the term 'upgraded' in *NE1* is used to reflect this.<sup>5</sup>

<sup>4</sup> It may be that these positions apply to other domains, but the focus here is on dynamical and mechanistic models in cognitive science.

<sup>5</sup> Here, the worry may arise that it is possible for all dynamical models to be upgraded in a way that makes them mechanistic, so that the distinction between mechanistic and non-mechanistic models collapses entirely. I address this worry at the end of Sect. 5.

Of course, one should differentiate here between two questions: whether it is possible for a model to be both dynamical and mechanistic, and whether there are in fact examples of models that are both. As they are phrased, the positions are answers to the latter question; a question which I believe should be answered affirmatively. Indeed, reasoning from the actual to the possible, one might suffice here with simply pointing out that some authors have convincingly shown some instances of dynamical models to be merely predictivist, and other instances to be both predictivist and mechanistic, as Zednik (2011) has done.<sup>6</sup> I agree with Zednik that “the differential equations and graphical representations that figure in many dynamical explanations can be, in principle as well as in practice, interpreted as representations of cognitive mechanisms” (ibid, p. 247). However, I think it is worthwhile not to take this point for granted, and to take the time to compare some of the more important notions from dynamical cognitive science with their mechanistic counterparts. This will allow us to understand more precisely what the claim that dynamical and mechanistic models show a partial overlap amounts to, what it is for a dynamical model to embrace mechanistic commitments, and hence to flesh out the distinction between mechanistic and non-mechanistic varieties of dynamical explanations. Together with the examples, discussed both here and elsewhere in the literature, of actual models that are both mechanistic and dynamical, I will take this to be sufficient for the truth of *NE1*.

I will proceed as follows. In the next Sect. 2.2, I will make a short inventory of the most important aspects of the *systems* studied by dynamical cognitive science, and a similar inventory of the most important features of *mechanisms*. In the succeeding Sect. 2.3, I will compare how both types of models are usually thought to *explain* phenomena. The combined material of these sections sets the stage for a full comparison between the dynamical and the mechanistic approach in Sect. 2.4. In Sect. 2.5 I will discuss (and reject) some potential objections to the non-exclusivity of dynamical and mechanistic explanations. Here, I will also make clear in what sense exactly dynamical explanations are predictivist. In Sect. 2.6 I will list and briefly discuss some positive reasons for non-exclusivity. Of course, having established that a model can be both dynamical

<sup>6</sup> It should be noted however, that Zednik himself would probably not agree with this reading, as he says that the examples of mechanistic dynamical models he considers “resemble mechanistic explanations *rather than* covering-law explanations” (2011, p. 245, my emphasis). As he also thinks that Kelso’s dynamical model of bimanual coordination is a predictivist (or in his terms, a CL) explanation (Ibid, p. 244), it seems that Zednik subscribes to position *E3* as defined above. However, in discussing Walmsley’s (2008) analysis of Thelen et al.’s (2001) dynamical field theory of infant perseverative reaching (a theory I shall briefly consider later in this article), he does admit that it can be used to “derive, by way of deductive inferences, predictions about goal-directed reaching in actual and counterfactual circumstances” (Zednik 2011, p. 250), although he claims that emphasizing this particular feature of the model over its mechanistic features is misleading (Ibid, p. 250). In my view however, regardless of what feature one chooses to emphasize, to vindicate the non-exclusivity thesis (and thus show that either position *NE1*, *NE2* or *NE3* must be correct) it is enough that both are present. The difference can be brought out clearly when we consider carefully what Zednik says of his examples: “In each case, the explanation proceeds by identifying the component parts and operations of a mechanism and by showing how the organized activity of these parts and operations produces the phenomenon being explained. *Therefore*, Thelen et al. and Beer each provide a counterexample to the received view of dynamical explanation: some dynamical explanations are mechanistic explanations rather than covering-law explanations” (Ibid, p. 255, my emphasis). As I endorse position *NE1*, on my account, Zednik’s conclusion does not follow.



and mechanistic does not give us the precise relation between these types of models, a relation which I shall endeavor to make clear in 2.7.

## 2.2 Dynamical systems and mechanisms

The conceptual and technical issues involved in dynamical models are often complex. Fortunately, dynamical cognitive science (and indeed dynamical systems theory generally) has received attention from philosophers since at least the early 1990s, so a lot of the analytical work has already been done, and done quite well. The precise details will vary between different accounts, but here is (with some slight adaptations) how van Gelder and Port (1995, pp. 5–17) propose to understand some of the more important concepts involved.

First, as should be evident from its name, dynamical systems theory studies *systems*, where systems are understood as sets of changing aspects of the world. Not any random collection of changing aspects constitutes a system however; the aspects must belong together in that the changes they undergo must depend on the changes in one or more of the other aspects. Besides this requirement of interaction, the collection of changing aspects must also be self-contained, in that any aspect that is responsible for a change somewhere in the system also belongs to that system. Second, such systems always exhibit a *state*, where a state is the way the aspects are at a particular time. Third, all systems have a *state space*, which is the totality of states the system can be in. Fourth, this state space allows one to interpret the *behavior* of the system, i.e. the changes in its state over time, as a series of points in its state space. Finally, a *dynamical system* is a system in which (i) the future behavior cannot depend on any state before the current state and (ii) there is some rule of evolution describing the behavior of the system as a function of its current state.

The core idea behind the dynamical approach to cognition is that the systems responsible for cognitive capacities, such as speech production, memory retrieval, pattern recognition etc., are dynamical systems. But of course, the examples just given are also precisely the kind of capacities that the mechanists would claim are realized by *mechanisms*, which, like dynamical systems, are also a special kind of system. Here, we see why it is helpful to flesh out the relation between dynamical systems and mechanisms. Both the dynamicists and the mechanists want to avoid triviality, and therefore recognize the need to put constraints on what counts as a relevant or interesting system. According to the dynamicist, not every system is a dynamical system, and likewise, according to the mechanist, not every system is a mechanism. Thus, we need to compare these sets of constraints. If they are incompatible, this might lead to what we may term ‘systemic overdetermination’, i.e. the idea that a given capacity like speech production may be realized by two different, mutually incompatible kinds of systems.

Here are the most important ingredients of mechanisms according to the influential ‘MDC account’ of mechanisms (Machamer et al. 2000). According to this account, mechanisms are “entities and activities organized such that they are productive of



regular changes from start or set-up to finish or termination conditions” (Ibid, p. 3).<sup>7</sup> Thus, mechanisms involve at least the following ingredients: *activities*, which are the producers of change, *entities*, which engage in these activities, and *organization*, i.e. the manner in which the activities and entities work together to produce the overall behavior of the mechanisms. In turn, this *behavior* is the explanandum phenomenon, or the capacity of a system that stands in need of explanation. Moreover, these aspects of mechanisms exhibit certain properties. Activities have temporal order, rate, and duration. Entities have location, structure, and orientation. Finally, a mechanisms’ organization ensures that its behavior exhibits regularity. This regularity concerns the similar way the mechanism proceeds to its end state, given the same start-up conditions—a phenomenon dubbed ‘productive continuity’ in the MDC account.

At this point, important similarities between dynamical systems and mechanisms start to emerge. Before tacking stock however, let us first address the explanatory side of things, and compare how both approaches are thought to explain.

### 2.3 How dynamical and mechanistic models explain

Dynamical cognitive science is an application of *dynamic systems theory*, which studies and describes the behavior of complex dynamic systems with the help of mathematical tools such as differential and difference equations. It also involves elements from computer science, most notably computer models and simulations, which can be used to describe all kinds of dynamic systems. Some of these systems can exhibit a property known as *complexity*, meaning they are not simply a collection of static entities that can be studied individually, but rather dynamic networks of entities and interactions between these entities. Dynamic systems theory then tries to describe how relationships between parts give rise to the collective behavior of (both natural and artificial) systems. Its applications are many, and it can be used to study systems found in diverse fields, such as artificial intelligence, economics, chemistry, geology and social psychology. As we have seen, such dynamical systems are described in terms of a so-called state space, and particular trajectories running through that space. In turn, these trajectories can be described using equations.

This mathematical character, together with the possibility to run computer simulations, makes dynamic systems theory apt for the purpose of prediction, and it is therefore often used to analyze systems that are of particular interest to human society. For example, in the case of the last two disciplines mentioned, geology and social psychology, dynamic systems theory is used to attempt to predict earthquakes and the behavior of crowds in confined spaces. Some tools used in dynamic systems theory to

<sup>7</sup> Like in the case of dynamical systems, the various accounts offered in the literature do vary somewhat in their details. To give just two examples, Bechtel and Abrahamsen (2005) define a mechanism as a “structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.” (p. 423). According to Craver (2007), a mechanism is “a set of entities and activities organized such that they exhibit the phenomenon to be explained (p. 5). Despite the differences, all these accounts include the ingredients of entities/component parts, activities/components operations, organization, and overall behavior/phenomenon. More importantly for our present purposes however, all view mechanisms as complex systems.

describe complex systems are also useful to study cognition, such as trajectories running through state space. With these tools, one can model behavior that is thought to be indicative of certain cognitive capacities. If dynamic systems theory is thus applied, we can speak of dynamical cognitive science.

Dynamical cognitive scientists try to explain a cognitive phenomenon by constructing a model of it. The explanandum (the cognitive phenomenon or activity to be explained), interpreted as a change in the relevant cognitive system, is modeled by an abstract dynamical system, which specifies variables as well as differential and difference equations governing the relations between those variables. These models are then used to derive predictions (in more complex models, such predictions have to be derived by means of computer simulations) that are subsequently matched up with experimental data. If the correspondence is sufficiently close, the model (the abstract dynamical system devised by the scientist) is taken to be similar to the dynamical cognitive systems producing the target phenomenon (ibid, p. 620).

So what about mechanistic explanation? The MDC account stresses that mechanisms are often represented by means of diagrams that specify the spatial and structural relations between the parts of a mechanism (Machamer et al. 2000, p. 8). Although MDC themselves do not use the word, from the subsequent literature it is clear that the proponents of mechanistic explanations view such representations as models (cf. Glennan 2005; Bechtel and Abrahamsen 2005; Craver 2006). According to Glennan, a model of a mechanism “consists of (i) a description of the mechanism’s behavior (...) and (ii) a description of the mechanism that accounts for that behavior” (2005, p. 446). Obviously, in the present context we are interested in the second part of this definition. A mechanistic explanation of a cognitive capacity or activity explains that capacity by describing, at a relevant level of detail, the mechanism responsible for it: its activities, entities, and organization. In short, whether conveyed verbally or by means of a diagram, the model should describe all the relevant ingredients of a mechanism, as described in Sect. 2.2.

## 2.4 Taking stock

We are now in a position to make a comparison between dynamical and mechanistic explanations, both with respect to the systems that are studied, and the way these systems are studied. Below is a table in which the most important aspects of both types of explanations are contrasted with each other.

As we can see, there is a lot of common ground here. Both dynamical and mechanistic explanations attempt to explain the behavior of some sort of special system (both are more than simple horizontal explanations, but go ‘in depth’), both track changes in that system between two points in time, and both explain by means of a model. Of course, we should not get overexcited about common terminology. ‘Model’ means something different in both cases. On the face of it, the biggest obstacle to a model being both dynamical and mechanistic is that, while the former use quantitative tools to describe the changes in its target-system as unfolding in real time, the latter analyze its target-system in terms of a discrete series of qualitatively described operations. In the next subsection, I will take a closer look at some of the more salient differences.

Dynamical explanation	Mechanistic explanation
<p><i>System</i> A set of changing aspects of the world that are (i) interactive and (ii) self-contained</p> <p><i>State of the system</i> The way the aspects belonging to a system happen to be at a particular time</p> <p><i>State space</i> The totality of overall states the system might be in</p> <p><i>Behavior of the system</i> The particular changes the aspects of a system undergo, expressible as a series of points in the state space. Also used to refer to the explanandum: the overall cognitive behavior or capacity exhibited by the system being modeled</p> <p><i>Dynamical system</i> A system in which: (i) the future behavior cannot depend on any state before the current state, (ii) there is some rule of evolution describing the behavior of the system as a function of its current state</p> <p><i>Model</i> A description of an abstract dynamical system, which specifies abstract variables as well as differential and difference equations governing the relations between those variables. It can be used to predict the explanandum behavior of the system</p>	<p><i>System</i> A set of changing aspects (entities, activities and their organization) of the world that are (i) interactive and (ii) self-contained<sup>a</sup></p> <p><i>State of the system</i> The way the aspects belonging to a system happen to be at a particular time<sup>b</sup></p> <p><i>State space</i> The totality of overall states the system might be in<sup>c</sup></p> <p><i>Behavior of the system</i> The transition of a system from one state to another. Also used to refer to the explanandum: the overall cognitive behavior or capacity exhibited by the system being modeled</p> <p><i>Mechanism</i> Entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. Thus, what distinguishes mechanisms from other, non-mechanistic systems is their regularity; i.e. that they exhibit productive continuity (Machamer et al. 2000, p. 3)</p> <p><i>Model</i> A description of the mechanism thought responsible for the explanandum; it describes how the entities, activities and organization conspire together to produce the overall behavior or capacity</p>

<sup>a</sup> It is unfortunate that the proponents of mechanistic explanations as a rule do not specify what they mean by 'system'. Nevertheless, talk of (complex) systems is abundant in the literature on mechanisms (Machamer et al. 2000; Glennan 2002, 2005; Bechtel and Abrahamsen 2005; Craver 2007). The notion was inherited from the work of Wimsatt (1976) and subsequently further developed by Bechtel and Richardson (1993) and Glennan (1996). Moreover, the mechanist must assume *some* implicit notion of it, since, as we already observed, they cannot hold that all systems are mechanisms, on pain of triviality. This means that it is up to the reader to flesh out a notion of system that is both consistent with this literature and makes mechanisms, understood along MDC-lines, a special subset of systems. In this regard, it seems that van Gelder and Port's notion of system can be imported into the mechanisms discussion without too much difficulty, if we take the 'changing aspects' to include entities, activities and their organization

<sup>b</sup> According to the MDC account (Machamer et al. pp. 11–12), two such states are special: the set-up conditions (in a mechanistic explanation, such conditions are typically described in an idealized fashion) and the 'privileged' termination conditions (often a state of rest or equilibrium). Of course, we need not follow the MDC account here, but the point is that, like dynamical models, a mechanistic model will begin with a description of one state, and end with another

<sup>c</sup> The term 'state space' does not feature in the literature about mechanistic explanations, but one can see how it might at some point be useful to talk about all the possible states a mechanism can be in

## 2.5 Overcoming some potential barriers

Unlike earlier attempts to understand cognition, such as classic computationalism or connectionism, which interpret time as a discrete sequence of steps, dynamical explanations consider cognitive capacities to unfold in real time (van Gelder and Port 1995, p. 2). For various practical reasons, time can be treated as a sequence of discrete steps, but the changes to the aspects of the system itself take place in real time. At first glance, this seems to constitute an important difference with mechanistic explanations,

which often confine themselves to tracking changes in the parts of the mechanism in a serial fashion. However, this restriction is not of a principled nature, and as [Bechtel and Abrahamsen \(2011\)](#) demonstrate with numerous examples, it is possible for an explanation identifying parts, operations and organization of a mechanism to be supplemented with the concepts and tools of dynamical explanation, so as to allow us to study the orchestrated behavior of mechanisms in real time.

Second, there is the issue of explanatory levels. It is uncontested that mechanistic models are often highly inter-level, crossing borders between many levels of descriptions and grain sizes. Dynamical models however, are often held to be intra-level ([van Gelder 1998](#)). Many of the paradigmatic examples of dynamical models, such as the Haken et al. model of rhythmic finger tapping ([Haken et al. 1985](#)) specify variables and parameters at the same level as the target-phenomenon. In fact, this presumed intra-level character of dynamical models is one of the reasons [Kaplan and Craver \(2011\)](#) argue that they merely (re)describe the explananda instead of explaining them. Yet not all dynamical models are strictly intra-level. To give a (by now) stock example, Thelen et al.'s (2001) model of infant perseverative reaching (see note 6) explains the A-not-B error by referring to activities and processes located at many levels, including behavioral (reaching), neural (motor planning) and psychological (memory) levels. As this example shows, there is no principled reason why the variables specified in a dynamical model cannot be taken to refer to entities located at different levels. Thus, the issue of explanatory levels is no barrier to dynamical models being upgraded to become mechanistic.

Third, although mechanistic explanations can be augmented with dynamical tools and concepts, often they are restricted to qualitative descriptions. On the other hand, it is undeniably the case that many dynamical models are quantitative, in that they take data, i.e. a series of measurements of the behavior, and capture these measurements with equations and the state space. If carried out successfully, the numerical sequences yielded by the equations and the state space match up to those exhibited by the target system, not only for already observed states, but also for future states. In this way, such models allow for precise quantitative predictions. However, when such precise measurements are impossible, dynamical models can still be used to give qualitative characterizations of cognitive phenomena ([van Gelder and Port, p. 16](#)). A system may exhibit dynamical properties such as oscillations, entropy, equilibrium states, etc. The solutions to the equations governing these properties may be intractable, which is what motivates researchers (starting with Poincaré) to conceptualize dynamical systems geometrically, i.e. as positions, distances and trajectories in a space of possible states. Even without exact solutions to the governing equations then, it is possible to put constraints on the kind of system that is being modeled. As such, a qualitative dynamical description of a system may provide preliminary understanding that can be used as a basis for further, quantitative investigation ([van Gelder 1998, p. 621](#)).

In short, the predictions made by dynamical models need not always be precise and quantitative. Nevertheless, they are inferential in nature. Here we come to the heart of the matter concerning the relation between CL and dynamical cognitive science. Although dynamical explanations presumably do not always adhere to all the strict requirements of adequacy Hempel and Oppenheim (1948, p. 137) proposed, they have important predictivist qualities. To return to our previous example, Thelen et al.'s

(2001) model of infant perseverative reaching infers the A-not-B error as a consequence of an equation relating the current state of the movement planning field, together with some boundary conditions (e.g. general and specific aspects of the reaching task). However, this inference is not deductive, since Thelen's model cannot take into account all contextual factors and experimental conditions (Gervais and Weber 2011, p. 38).

In general, the rules which govern the behavior of dynamical systems need not be interpreted as universal generalizations, but can also be thought of as default rules (Ibid, pp. 37–38). What matters at this point is that the explanations furnished by dynamical models are inferential in character, and as such that they can be used to make predictions (although it remains to be seen whether this is particular strategy is genuinely explanatory or not). Likewise, mechanistic models can be used to make qualitative predictions about future behavior of mechanisms (and thus have inferential credentials)—again, these predictions will often be imprecise, as there is often no way to know, let alone eliminate, all the relevant variables. Thus, as in the case of real versus sequential time, the qualitative/quantitative difference between dynamical and mechanistic explanations is real and important, but it does not constitute a principled argument against the possibility of a dynamical model being upgraded to also be mechanistic.

## 2.6 Some positive considerations

Besides defusing these potential stumbling blocks however, there are also some positive reasons for thinking that a dynamical model can be upgraded to be mechanistic as well. We have encountered some of these points already in one way or another, so I will confine myself to a brief summary.

- (1) *Reciprocal causation* Dynamic explanations allow for two or more factors of the system to simultaneously influence each other. Likewise, mechanistic explanations can accommodate causal feedback loops. For example, in the case of the circadian rhythms in mammals, entrainment of the suprachiasmatic nucleus in accordance with the external cues is managed by hormonal (e.g. melatonin) feedback loops involving the pineal gland (cf. Moore 1997).
- (2) *Embeddedness* Both mechanisms and dynamical systems continuously influence, and are continuously influenced by, their surroundings. This is compatible with embodied/embedded approach, according to which cognition is the smooth, continuous interplay between mind, body and environment (Clark 1997; Wheeler 2005).
- (3) *Communication* Mechanists often point out that communicating inferential explanations usually involves linguistic or mathematical representations, whereas mechanistic explanations are often represented by diagrams (Bechtel and Abrahamsen 2005; Bechtel 2009, p. 553). However, despite being inferential/predictivist, dynamical explanations lend themselves very well for graphical representations. Non-linear equations, state spaces, attractor landscapes, vector fields, etc., can all be represented by means of diagrams.
- (4) *Philosophy* Finally, there is also an important philosophical similarity between dynamical and mechanistic explanations of cognitive phenomena. They are both

motivated by a philosophical conviction, in that they make certain claims about the nature of the systems they are studying. In dynamical cognitive science, this is known as the *dynamical approach*. Adherents of this approach do not only make the methodological claim that the tools of dynamical systems theory are useful in describing and explaining natural cognitive systems, but they also endorse the ontological claim that cognitive systems *are* dynamical systems (van Gelder and Port 1995, p. 5). Likewise, the assumption of the mechanistically minded philosopher of cognitive science is that cognitive systems *are* mechanisms.

## 2.7 Intermediary conclusion: dynamical models are predictivist, and can be upgraded to become mechanistic as well

I take the preceding comparison between dynamical and mechanistic models to have established two points. First, for better or for worse, dynamical models are predictivist in that they infer their explananda from some generality plus boundary conditions. Second, dynamical and mechanistic models overlap to a certain degree: it is possible for a dynamical model to become mechanistic by embracing certain commitments about the system it describes (see below). But of course, some might object that this still does not vindicate position *NE1* as defined in Sect. 2.1: one could insist that *pure* dynamical models are necessarily wholly mathematical in character, dismissing the examples of mechanistic dynamical explanations usually offered in the literature as no longer truly dynamical. This is the strategy followed by Fresco (2012, pp. 371–373), who, in discussing one of Zednik's (2011) examples of mechanistic dynamical explanations (Randall Beer's model of perceptual categorization in a simulated brain–body–world system), agrees that it is mechanistic, but maintains that “dynamicism proper is a non-mechanistic explanatory framework” (Fresco 2012, p. 372).

I believe we should not be rigid about the terminology here. If one stipulates that dynamical models only consists of applying mathematical formalisms to evolving systems, then of course a mechanistic dynamical model is no longer a purely dynamical model, but this misses the point, namely that a dynamical explanation can be supplemented with mechanistic information, and that doing so can provide additional insight into the system under scrutiny. This point is not trivial, given the fact that some authors believe that the dynamical and the mechanistic approach are mutually exclusive alternatives. Fresco is right however, in his assertion that we need a distinguishing feature of mechanistic instances of dynamical explanations that sets them apart from non-mechanistic instances, if we want to avoid setting foot on a slippery slope that ultimately renders all dynamical explanations mechanistic (Ibid, p. 371).

I think that the most natural way to think about this issue is to view mechanistic explanations of cognitive capacities as *interpreted* dynamical explanations: the variables and parameters of a dynamical model are taken to refer to certain aspects of the cognitive system responsible for the explanandum behavior. Of course, as should be evident from the comparison above, in dynamical cognitive science such interpretations are already present: cognitive systems are held to be real, dynamical systems underlying real behavior, and the differential equations are meant to describe real behaviors such as infant perseverative reaching—hence, the slippery slope worry advanced by Fresco is real. In general, whether a dynamical explanation is also mech-

anistic depends on how extensive this interpretation is. This may sound vague, but fortunately, the literature offers us a handle on these matters. Consider Kaplan and Craver's so-called 3M constraint:

**(3M)** In successful explanatory models in cognitive and systems neuroscience (a) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism (Kaplan and Craver 2011, p. 611).

Although Kaplan and Craver are concerned with dynamical models in systems neuroscience rather than in cognitive science generally, it is easy to see that it can also be applied in the present context: a dynamical model can be upgraded to become mechanistic as well, by meeting the requirements set out by Kaplan and Craver in 3M. As there are examples of actual dynamical models that are upgraded to meet these requirements (cf. Kaplan and Bechtel 2011, pp. 441–442), I take the truth of NE1 to have been established.

### 3 Why predictive power, counterfactual support, and unification are not sufficient

So far, I have argued for the first claim of the introduction: that there are mechanistic and non-mechanistic varieties of dynamical models. Let us now turn to the second issue: explanatory power. Recall that the mere description worry is about the *non-mechanistic* variety of dynamical explanations; as I have said in the introduction, nobody denies the explanatory power of mechanistic dynamical explanations. The point is that we need to be clear that an example is genuinely non-mechanistic, and *then* assess its explanatory power. The question then turns on whether there exist dynamical models that do not meet 3M, yet are still explanatory. In this section, I shall first consider—and reject—the reasons usually offered in the literature to argue for this view. After this, I will present scientific understanding as a fourth argument (Sect. 4) and illustrate my claims with the Voss strong anticipation model (Sect. 5).

So why do some authors maintain that non-mechanistic dynamical models explain? They offer three related arguments, each one building on the previous. Let us review each of them in turn.

The first argument is that dynamical models allow for testable predictions—as we have seen, this view has been labeled ‘predictivism’. This argument for the explanatory power of dynamical models was already hinted at by van Gelder (1998, p. 625),<sup>8</sup> but

<sup>8</sup> Van Gelder also offers another argument (1998, p. 625), namely that dynamical systems theory is also used to study systems in many other domains besides cognitive science, so that rejecting dynamical models of cognitive systems as non-explanatory would have the consequence of rendering all these models non-explanatory too. I will not consider this argument in detail, but as Walmsley (2008, p. 337) rightly points out, being a *reductio*, it offers no help against someone willing to deny the explanatory power of dynamical models in general.



has subsequently been raised by others (Chemero and Silberstein 2008; Walmsley 2008; Stepp et al. 2011). However, it can easily be countered. Indeed, it is precisely the possibility of predicting future states of events from phenomenological laws that undermines the most famous predictivist account of all, namely Hempel's CL model (Salmon 1989). We can predict a storm by observing a drop in a barometer, but this does not explain the storm; we can predict how high the tide will come by looking at the debris left by the previous tide, but this does not explain the height of the tide, etc.

Simply predicting actual behavior then is not enough. But what about predicting counterfactual behavior? To be sure, describing the behavior of a dynamical system as a state space does allow one to answer a number of what-if-things-had-been-different questions (Walmsley 2008; Stepp et al. 2011). But again, this does not suffice: the mechanist will maintain that in order to explain, one must show *why* such counterfactual predictions follow. In sum, a law may provide predictions of actual or counterfactual states of affairs, but that does not explain why the law applies in the first place. Indeed, far from being an explanans, the law is itself in need of explanation.

Finally, the predictivist line of thinking can be augmented into a third argument: *unification*. The idea is as follows. We can use counterfactual predictions provided by the dynamical system to suggest further experimentation (Chemero 2009), such that a model originally intended to study one phenomenon can be used again, with slight changes, to study another phenomenon. Thus, it brings together diverse phenomena within the scope of a small group of closely related models, which is taken to indicate that such models can provide unification in the Friedman/Kitcher sense (Stepp et al., pp. 432–435).

But again, this is not sufficient for explanation. Of course, unification is an important goal in science, and applying a general model to disparate phenomena to reveal underlying behavioral patterns across systems is valuable in itself, but that is not what is at issue here. Rather, the question is whether a model explains the behavior of the system it describes, *because* its description also applies to the behavior of other systems. Without further argument, this does not follow (Kaplan and Bechtel 2011, p. 441).

Thus, the arguments traditionally offered to uphold the explanatory power of non-mechanistic dynamical explanation do not stand up to scrutiny. In the remainder of the paper, I will try to show that there is an additional argument that does, namely that such models can provide scientific understanding. First however, we must establish what we mean by this term.

## 4 Scientific understanding

While the topic of explanation has been extensively discussed in the philosophical literature, understanding has received far less attention. This lacuna may well be a logical empiricist inheritance. For Hempel (1965, pp. 425–426), understanding is merely a pragmatic, psychological byproduct of explanation, and as such is epistemically irrelevant. More recently, Trout has argued that those properties of an explanation that are relevant to evaluating its merits pertain to features of external objects that exist independently of the psychology of the scientists using these explanations (2007, p. 217).

On this view, understanding, conceived as a kind of subjective confidence that an explanation can bestow on the scientist using or grasping it, has no place in (to use Reichenbach's terminology) the context of justification, though it might play a role in the context of discovery. Indeed, experimental research in cognitive science suggests that such subjective feelings can even be dangerous, as they are liable to be influenced by overconfidence and hindsight bias (Trout 2002, pp. 223–229). Addressing these objections to the epistemic relevance of understanding is all the more important in the present context, since many prominent mechanists favor an ontic, objectivist account of explanation (Machamer et al. 2000; Glennan 2002, 2005; Craver 2007). Underlying this view on explanation is a broadly ontic account of causation along the lines set out by Salmon (1984).<sup>9</sup>

In any case, the gist of these challenges is that understanding, conceived as a subjective, 'aha' feeling produced in a scientist as he or she grasps a theory, is neither necessary nor sufficient for explanation; and that it does not offer us a criterion by which to judge the correctness of an explanation either.

To be sure, in so far as understanding refers to the persons involved in providing or grasping the explanation, it is not an objective notion. Yet from this it does not follow that it amounts to nothing more than an entirely subjective feeling (of which I would agree with Hempel and Trout that it is neither necessary nor sufficient for explanation). For the purposes of my argument, I will focus on a specific type of understanding, as described by the contextual theory of scientific understanding developed by De Regt and Dieks (2005) and later expanded by De Regt (2009). Unlike the subjective notion of understanding (the 'aha' feeling), which is an effect of an explanation on the explainer, this notion of understanding is pragmatic in that it focuses on how a scientific theory can be *used*. I will argue that, unlike the subjective notion, this pragmatic notion is necessary for explanation, and for distinguishing good from bad explanations.

The basic observation underlying the contextual theory of understanding is that the specific skills or patterns of use that are enabled by a given scientific theory are necessary to take into account when evaluating its explanatory value.<sup>10</sup> For example, consider Hempel-style, inferential explanations. Here, simply knowing the laws or general principles governing a particular event, plus the relevant background conditions, is not enough to explain this particular event. Something more is required, namely the *skills to derive* the particular event from the laws plus background conditions. Although when explicated, the resulting explanation is a formal proof in which each step is justified by appealing to explicit inference rules, cognitively speaking the process of constructing such a proof is not rule-governed. Instead, deciding what inference rule to invoke, is an acquired skill (Brown 2000). Moreover, this holds not only in the search or new proofs, but also in the evaluating of a given proof (Brown 1988,

<sup>9</sup> It needs to be said though, that this ontic conception of explanation has not gone unchallenged, even within the community of philosophers who concern themselves with mechanistic explanation. In particular Bechtel (2008), along with Wright (2012) has rejected such an ontic reading in favor of an epistemic one (see Illari 2013 for an reconciliatory overview of the arguments on both sides). Nevertheless, even on the epistemic reading, the objections still stand, and 'understanding' requires some measure of success in understanding the world, rather than simply a feeling of satisfaction (Waskan 2011).

<sup>10</sup> In claiming that these skills are necessary, I move somewhat beyond de Regt, who uses the term 'crucial' instead (2009, p. 588).

pp. 170–171). Thus, in constructing and evaluating the kind of inferential explanations Hempel advocates, it is necessary to take these skills into account.

The above applies even more to model explanations. Instead of relying on a deductive relation between explanans and explanandum, models have a more complex and vague relation to their target-phenomena, involving abstraction, approximation and idealization. No formal rules are invoked in constructing models, and scientists rely instead on rules of thumb and instrumental values (Cartwright 1983). According to the Regt “This implies that the construction of a model, as well as the assessment of the model as a good (or good enough) representation of the system, is a process in which scientists have to make pragmatic decisions and must accordingly rely on skills and judgment” (2009, p. 591).

The upshot then is that there is a sense of understanding that, although not objective, is not entirely subjective either, and that this sense of understanding is a necessary condition on adequate explanation. So far however, I have not yet clarified what is meant by ‘using’ a theory.

At the core of the contextual theory of understanding are two criteria (De Regt and Dieks 2005, pp. 150–151):

**CUP** (criterion for understanding phenomena): A phenomenon  $P$  can be understood if a theory  $T$  of  $P$  exists that is intelligible (and meets the usual logical, methodological and empirical requirements) (Ibid p. 150)

**CIT** (criterion for the intelligibility of theories): A scientific theory  $T$  is intelligible for scientists (in context  $C$ ) if they can recognize qualitatively characteristic consequences of  $T$  without performing exact calculations (Ibid p. 151)

The first criterion CUP says that in order to understand a phenomenon, it is necessary to have a theory, or in our case a model of that phenomenon, that is intelligible. Note that this is true even for predictivists: in order to make predictions with a theory, it is necessary that the theory is intelligible. Of course, this raises the question when a theory is intelligible. The second criterion CIT says that we must be able to recognize qualitatively characteristic consequences of  $T$  without necessarily possessing the skills to perform exact calculations with it.

It is important to note the pragmatic character of CIT (and hence of CUP). It refers to both the scientists who use a theory, and the context in which they use it. As the context will vary according to the skills, capacities and background knowledge of the scientists in question, CIT allows a theory to be unintelligible to some, while intelligible to others. Although intelligibility is a context dependent value ascribed by scientists to theories (rather than an intrinsic property of theories), this does not mean that it is simply a matter of taste. Skills, capacities and background knowledge are shared within scientific communities (Kuhn 1970), so that while there are no universal criteria for intelligibility, within a context CIT can nevertheless be used as a test to determine whether a theory is intelligible. Thus, despite being pragmatic, the notion of understanding considered here is not merely a subjective feeling on the part of the explainer, so that Trout’s objections do not apply.

De Regt and Dieks illustrate their theory with the explanation of Boyle’s gas laws by the kinetic theory of gas. The idea is as follows. Boyle’s laws make a number of assertions about the behavior of gas in closed spaces. Specifically, it states that if we

increase the temperature of a gas, the pressure it exerts on the walls of its container will increase. Now the kinetic theory of gas can give us a qualitative understanding of this behavior, without allowing us to make exact calculations. Knowing that gas is constituted by a cloud of freely moving molecules that accelerate if the temperature increases, allows us to understand why the pressure on the walls of the container increases: the molecules bounce against the walls at greater speeds. To make precise predictions (e.g. to say how much the pressure will increase when the temperature is increased by one degree), one would need to invoke the sort of mathematical equations that Boyle's law provides, but to get a general idea why this particular correlation exist, a qualitative understanding of the kinetic theory of gas is enough. According to the contextual theory of understanding then, if we recognize these qualitative consequences, the kinetic theory of gas is intelligible, and offers us understanding of the behavior of gasses in closed spaces.<sup>11</sup>

Two objections might be raised here. First, it may seem that this theory of understanding only serves to strengthen the case of the mechanists. After all, in the example of Boyle's gas laws, the qualitative understanding (namely why gas and temperature are correlated) is only gained after we have acquired knowledge about certain mechanistic details: namely, that the parts of a cloud of gas include entities (molecules) that engage in activities (moving freely) etc. Yet, as I will illustrate below with a further example, understanding can also be provided by a theory that is only hypothesized to be correct. A second objection might be that for the kinetic theory to afford understanding, some kind of overarching quantitative framework should already be in place, whether it is actively being used or not. Assuming that a theory has to be correct to offer understanding,<sup>12</sup> can we really confirm its correctness without making precise quantitative predictions? Although I agree that ultimately, to verify a model or theory, one needs to devise intervention experiments in order to test predictions, it does not follow that these interventions and predictions need to refer to exact quantities. In the case of the Boyle–Charles law, a statement like 'increasing the temperature will result in an increase in pressure' expresses a testable prediction without specifying any values.

The following example might help to bring these issues to the fore. When in the 1950s Woolley and Shaw first proposed that certain mental disorders were linked to serotonin deficiencies in the brain, they immediately recognized some qualitative consequences:

...it is the lack of serotonin which is the cause of the disorder. If now a deficiency of serotonin in the central nervous system were to result from metabolic rather than from pharmacologically induced disturbances, these same mental aberrations would be expected to become manifest. Perhaps such a deficiency is responsible for the natural occurrence of the diseases [...] If the hypothesis about serotonin

<sup>11</sup> As this example shows, CIT is not just a weaker form of predictivism. Although it is concerned with recognizing consequences of a theory or model, the focus is on the ability to use the theory or model, not on a Hempelian notion of expectability (that the explanandum 'was to be expected').

<sup>12</sup> I use the term 'correct' instead of 'true' to allow for models containing idealizations (and hence are literally false) to be explanatory (cf. [Strevens 2013](#)).

deficiency is accepted, then the obvious thing to do is to treat patients having appropriate mental disorders with serotonin (Woolley and Shaw 1954, p. 230).

Like in the case of the gas in the container, the reconceptualization of mental disorders as metabolic hormone deficiencies led the researchers to recognize some qualitative consequences, namely that administering the hormone will alleviate these disorders, barring the existence of some hidden common cause. Yet the reconceptualization itself does not yield precise predictions. For example, solely on the basis of Woolley and Shaw's proposal, it is impossible to calculate precisely how high a dosage should be in a given situation. Moreover, the consequences of an intervention will remain imprecise (for example, researchers will judge and rank the behavior of patients, or patients will fill out a questionnaire, both of which are subjective methods). Nevertheless, even in the absence of some overarching quantitative framework, the hypothesized dependency can be (and indeed, has been) confirmed by performing the experiment proposed by Woolley and Shaw. Thus, even though at the time it was still an unconfirmed hypothesis, and even though it did not allow us to make exact predictions, nevertheless, for Woolley and Shaw, their theory was intelligible, and offered understanding in the sense of CUP.<sup>13</sup>

This is a far cry from the insistence that the variables and their interdependencies correspond to the actual components, entities, operations and causal relations of the underlying mechanism is *necessary*, as 3M would have it. In effect, the contextual theory *decouples* understanding from the ability to perform exact calculations. As we have seen in Sect. 2.5, some dynamical models offer qualitative rather than quantitative insight into their target phenomena. Here, CIT implies that it is possible to recognize qualitatively characteristic consequences of a dynamical model, without being able to make the calculations needed to arrive at precise predictions. A dynamical system can be conceptualized in terms of trajectories to the space of possible states even in the absence of closed-form solutions to the governing dynamical equations. Even if we lack precise solutions to the mathematical equations capturing some sort of cognitive behavior, or if the equations themselves cannot (yet) be laid down, dynamical models can still provide us with a conceptual apparatus that provides us with understanding of that behavior. As van Gelder and Port state "...in the absence of a precise mathematical model, the language of dynamics can be used to develop qualitative dynamical descriptions of phenomena" (1995, p. 16). Applying this to Thelen's (1995) dynamical approach to the development of embodied cognition in infants, they continue (van Gelder and Port 1995, p. 17):

...Thelen [...] is concerned with understanding the development [...] of basic motor skills such as reaching out for an object. At this stage, no satisfactory mathematical model of this developmental process is available. Indeed, it is still a major problem to write down equations describing just the basic movements themselves! Nevertheless, adopting a dynamical perspective can make possible descriptions which cumulatively amount to a whole new way of understanding

<sup>13</sup> One might object that even in the hypothetical case, the understanding is only provided by the reference to mechanistic details, not by the regularities—I shall consider this objection at the end of Sect. 5.

how motor skills can emerge and change, and how the long-term developmental process is interdependent with the actual exercise of the developing skills themselves. From this perspective, particular actions are conceptualized as attractors in a space of possible bodily movements, and the development of bodily skills is the emergence, and change in nature, of these attractors over time under the influence of factors such as bodily growth and the practice of the action itself.

Thus, even if the equations governing a dynamical system do not admit of solutions, or even if for some reason (e.g. complexity) the equations cannot (yet) be laid down, so that we cannot derive precise predictions, adopting the dynamical approach it is still possible to recognize qualitatively characteristic consequences of a model, so that it is intelligible along the lines of CIT and offers scientific understanding.

In the next section I will argue that a dynamical model can fail the 3M constraint, yet still provide understanding in the sense outlined above, by considering Voss's strong anticipation model.

## 5 Voss' strong anticipation model

Stepp et al. appeal to Voss's strong anticipation model (2000) to explain certain features of the mammalian circadian system, i.e. the ability of mammals to maintain an endogenous oscillation of about 24 h. Although this ability is endogenous, in maintaining it the responsible system uses external cues, most notably light (the 'Zeitgeber') to entrain the various neural subsystems in a process of continual readjustment. Now this functional analysis, although still very rough, already yields some more specific tasks that need to be achieved, such as the passing of information from the visual system to the particular brain structures involved (notably the suprachiasmatic nucleus or SN), feedback loops needed for the purpose of entrainment, etc. Stepp et al. (2011, p. 428) consider a number of features of circadian rhythms that need to be explained;<sup>14</sup> here, I will focus on the ability of a system to anticipate the light–dark cycle at the scale of the organism, organ, and cells. Anticipation here requires that one or more subsystems located at different levels follow the future trajectory of another, leading system (the circadian rhythm).

Strong anticipation can be used to attempt to explain this ability without invoking representations of the time of day, contrary to weak anticipation, which uses internal simulations of the light–dark cycle (Dubois 2003). In this way, strong anticipation does not offer the ability to make explicit predictions, relying instead on endogenous lawfulness of the system. In particular, the Voss model is used to show how the system synchronizes multiple oscillators in the SN, when the neurons that are entrained only indirectly by exposure to light are phase advanced to those that are directly entrained. In such a case, the anticipation of this synchronization can be expressed by a class of 'delay differential equations'. One such equation assigns functions  $f$  and  $g$  for the intrinsic dynamics of the oscillators  $x$  and  $y$  respectively, and relates these functions

<sup>14</sup> Stepp et al. introduce these features specifically as challenges to representational accounts of circadian rhythms (2011, p. 428), but that need not concern us here.

by means of coupling strength  $k$  and delayed feedback  $y_\tau = y(t - \tau)$ , such that (Ibid, p. 428):

$$\begin{aligned}\dot{x} &= f(x) \\ \dot{y} &= g(y) + k(x - y_\tau)\end{aligned}$$

In the system described by this equation, the driven system  $y$  has the propensity to anticipate the driver system  $x$ . Because in this system,  $x$  and  $y$  are in a unidirectional coupling arrangement, and the amount of anticipation is a function of the delay, many driven systems can be coupled to the same driver. This means that any differences between the intrinsic delay of  $y$  and other driven systems, result in different phase shifts with respect to the driver.

The next step is to translate the Voss model to circadian systems in mammals. Recall that the feature in need of explanation is the ability of a system to anticipate the light–dark cycle at the scale of the organism, organ, and cells. Just like the driven systems in the Voss model are coupled to a single driving system, in the mammals, many subsystems residing on many different levels are coupled to the same circadian system. Body temperature (organism), hormone production (organs), and cell regeneration (cells) are all regulated by the circadian clock. The point is that these subsystems have different intrinsic feedback delays, and as such need to anticipate the light day cycle—it is this that the strong anticipation model is supposed to explain. As we have seen, it does so in a manner typical of dynamical explanations: by describing the dependencies between these different systems in terms of differential equations.

Is this dynamical model mechanistic? Let us evaluate it according to the standards of 3M. First, the variables in the system correspond to what are called driven and driving systems, coupling strength, delayed feedback, etc. Some activities here, some properties to be sure, but no entities, except of course driven and driving *systems*. Yet if left unspecified, these act as mere placeholders: they are little more than a formal acknowledgment that there must be some entity engaged in the activities referred to by the model. Even if we translate them to the mammalian circadian system, as we have done above, so that the driven systems stand for subsystems such as body temperature and hormone production, still they denote abstract systems, not concrete entities. Next, do the dependencies between the variables correspond to the causal relations among the components of the mechanism responsible for anticipation? No, they correspond only to a unidirectional coupling arrangement between two oscillatory systems. If, again, we translate the Voss model to the mammalian system, then this unidirectional coupling arrangement simply reflects the acknowledgement that subsystems such as body temperature and hormone production are somehow linked with the circadian clock in such a way as to be able to anticipate the light–dark cycle. Is this link causal? It is tempting to interpret it this way, but neither the Voss model, nor its application to the mammalian circadian rhythm, warrants this inference.

Besides, the application of the model to the system responsible for the circadian rhythms of mammals is precisely this: an application. Kaplan and Bechtel are clear to point this out: “without specifying the parts and operations in the mammalian circadian system that performs such functions as coupling between oscillators, this account is



empty. The model remains only a *how possibly* model” (Kaplan and Bechtel 2011, p. 443, original italics). To be sure, one can construct a mathematical model that ‘saves the phenomena’, but in order to be mechanistic, one needs to embrace commitments about the actual components, parts, operations and causal relations of the target system itself. Clearly, the Voss model does not meet 3M.

There is little doubt then, that the Voss model is an example of a non-mechanistic dynamical explanation of a cognitive capacity. Indeed, Stepp et al. advance it as such, and this is affirmed by Kaplan and Bechtel (2011, p. 443). Thus, both parties agree that the Voss model as it stands is non-mechanistic, but differ on whether it is explanatory or not (Kaplan and Bechtel rejecting of course the three usual arguments discussed in Sect. 3), which makes it a suitable example for our purposes here.

Yet although it is non-mechanistic, the model does provide understanding of the phenomenon that it is supposed to explain, without necessarily enabling us to make precise predictions. Recall the explanandum: given that different subsystems of the mammalian circadian system, located on different levels, have different intrinsic feedback delays, they must somehow anticipate the light–dark cycle. So how do they accomplish this? Anticipation requires that one system follows the future trajectory of another system. Conceptualizing this accomplishment in terms of strong anticipation, allows us to recognize some qualitative consequences.

For example, as we have already seen, the coupling between the driving system  $x$  and the driven system  $y$  is unidirectional (as expressed in the equation above), so it follows that an arbitrary number of driven systems can be coupled to the same driving system. Furthermore, the dynamics of driving and driven system can also differ (expressed as  $f$  and  $g$  respectively in the equation). This means that a particular coupling dynamics can start out with one driving system that is subsequently replaced with another driving system: the driving–driven coupling dynamics can evolve such that the system switches to a second driving system. The driven system first anticipates synchronization with the first driving system. After the switch, the anticipated synchronization of the same driven system with the second driving system emerges gradually (Stepp and Turvey 2010). Although by itself, the model does not allow us to give precise calculations as to how gradually this happens, the gradual emergence of the anticipation of synchronization to the second driving system as such is a qualitative consequence of the delayed feedback loop. In the same way, when we apply this to the mammalian circadian system, we can understand why a change in the external cues, i.e. a change in the light–dark cycle, results in a corresponding shift in the circadian rhythm. Like the driven system in the strong anticipation models, the particular subsystems such as cell regeneration and bodily temperature first anticipate synchronization with the first circadian rhythm, only to gradually anticipate synchronization with the second rhythm after the switch. As such, it allows us to understand, in a qualitative way, the empirical observation that the shift in a specific biological circadian rhythm following a shift in the light–dark cycle, happens over time rather than instantaneously (Moore-Ede 1986).

As a second example, consider that strong anticipation shifts the burden of anticipation from the components of the system to the overall organization in which these components feature, allowing individual components to be ‘ignorant’. Thus, anticipation emerges from the natural unfolding of events, rather than from internal represen-

tations of the future. In turn, this implies that such systems are reactive, in the sense that the states of the system are a function of previous, not future states. The fact that the states of the driven systems correctly respond to future states of the driving systems purely arises from the organization of the system (Stepp and Turvey 2010). This means that applying the strong anticipation perspective to mammalian circadian systems has the qualitative consequence that subsystems such as heart rate and body temperature are phase advanced to the circadian clock purely by means of the organization of the circadian system as such, without any ‘intelligence loans’ (Dennett 1978) from the overall system to its components.

As Kaplan and Bechtel stress (2011, p. 443), the Voss model is only a how-possible model: it has yet to be verified if the mammalian circadian system works that way. But, just as in the case of Woolley and Shaw’s proposal, this hypothetical character has no bearing on the fact that the model conveys understanding in the sense of the contextual theory outlined in the previous section. In the given case, it is not the *actual* identification with parts, operations etc. that provides the understanding (since both parties agree this has not yet been accomplished), but the realization that the mammalian circadian system *might* operate in accordance with the Voss model.

Of course, one might object that even in the hypothetical case, it is the mechanistic features referred to in the dynamical hypothesis that provide the understanding, not the generalities involved, nor its mathematical character. There are two replies to this objection. First, it does not square with the literature. In their assessment of explanatory power in models, the mechanists propose requirements that go well beyond hypothetical models. They talk about ‘embracing commitments’ about the mechanism (Kaplan and Craver 2011, p. 603), or about the need to ‘identify’ the mechanism (Kaplan and Bechtel 2011, p. 441). These requirements, as described in the 3M constraint, are not met. In the case of the Woolley–Shaw hypothesis, and in the case of the Voss model, there has been no identification or localization of parts, components, causal relations, etc. (at least at the time they were proposed). Again, the reason I chose the Voss model as an example is precisely because both parties agree that it is non-mechanistic.

More importantly however, the variables and dependencies among the variables of the kind of models we have been discussing are *always* deemed to describe *some* system. After all, we are talking about dynamical *cognitive* science, which is the application of the mathematical tools of dynamical systems theory to cognitive phenomena: dynamical models are not just exercises in mathematics. The mere description worry is not leveled against mathematical equations as such, but against mathematical equations vis-à-vis some explanandum phenomenon. If we are going to count hypothetical dynamical explanations such as the Voss model as mechanistic, we run the risk of collapsing the distinction between mechanistic and non-mechanistic models entirely (and indeed, in such a case the mere description worry would not be so worrying). So it comes down to the question *how much* mechanistic commitment is needed to count a dynamical model as mechanistic. Here, I think it is reasonable to insist on the localization and positive identification of at least some parts—a requirement that is not met in the case of the Voss model.

## Conclusion

It is not denied by the mechanists that dynamical systems theory is making serious headway into cognitive science. Nor is it contested that dynamical cognitive science provides a new and conceptually innovative alternative to more traditional approaches to cognitive phenomena, such as computationalism and connectionism. What is disputed however, is that dynamical models can be explanatory without also being mechanistic. In this paper, I have taken issue with this view.

First, I argued that dynamical models can also be mechanistic, by embracing certain commitments along the lines of the 3M constraint. Second, I agreed with the mechanists in their rejection of the usual considerations put forward to argue for the explanatory power of non-mechanistic dynamical models. So I have explored a new one: understanding. Of course, although I hope to have shown that understanding in the sense explicated in Sect. 4 is necessary for constructing and evaluating explanations, one may dispute that it is sufficient for explanation.<sup>15</sup> In any case, I do believe that the argument from understanding presents a far more promising route than considerations about predictive power and unification. At the very least, it shows that non-mechanistic dynamical models go well beyond ‘mere descriptions’.

## References

- Bechtel, W. (1998). Representations and cognitive explanations. Assessing the dynamicist challenge in cognitive science. *Cognitive Science*, 22, 295–317.
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Oxford: Routledge.
- Bechtel, W. (2009). Constructing a philosophy of science of cognitive science. *Topics in Cognitive Science*, 1, 548–569.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421–441.
- Bechtel, W., & Abrahamsen, A. (2011). Complex biological mechanisms: Cyclic, oscillatory, and autonomous. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 257–285). New York: Elsevier.
- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.
- Brown, H. I. (1988). *Rationality*. London: Routledge.
- Brown, H. I. (2000). Judgment, role in science. In W. H. Newton-Smith (Ed.), *A companion to the philosophy of science* (pp. 194–202). Oxford: Blackwell.
- Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Clarendon.
- Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
- Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of Science*, 75, 1–27.
- Clark, A. (1997). *Being there: Putting brain, mind, body and world together again*. Cambridge, MA: Bradford/MIT Press.
- Craver, C. F. (2006). When mechanistic models explain. *Synthese*, 153, 355–376.
- Craver, C. F. (2007). *Explaining the brain*. Oxford: Oxford University Press.
- Dennett, D. C. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.
- De Regt, H. (2009). The epistemic value of understanding. *Philosophy of Science*, 76, 585–597.
- De Regt, H. (2013). Understanding and explanation: Living apart together? *Studies in History and Philosophy of Science Part A*, 44(special section), 505–509.

<sup>15</sup> See De Regt 2013 for various views pro and contra.

- De Regt, H., & Dieks, D. (2005). A contextual approach to scientific understanding. *Synthese*, 144, 137–170.
- Dietrich, E., & Markman, A. (2001). Dynamical description versus dynamical modeling. Reply to Chemero. *Trends in Cognitive Sciences*, 5, 332.
- Dubois, D. (2003). Mathematical foundations of discrete and functional systems with strong and weak anticipations. In *Lecture Notes in Computer Science*, vol. 2684, (pp. 110–132).
- Eliasmith, C. (1996). The third contender: A critical examination of the dynamicist theory of cognition. *Philosophical Psychology*, 9, 441–463.
- Fresco, N. (2012). The explanatory role of computation in cognitive science. *Minds and Machines*, 22, 353–380.
- Gervais, R., & Weber, E. (2011). The covering law model applied to dynamical cognitive science: A comment on Joel Walmsley. *Minds and Machines*, 21, 33–39.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49–71.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(Suppl.), S342–S353.
- Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences*, C, 36, 443–464.
- Haken, H. L., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movement. *Biological Cybernetics*, 51, 347–356.
- Hempel, C. G. (1965). *Aspects of scientific explanations and other essays in the philosophy of science*. New York: Free Press.
- Hempel, C. G., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15, 135–175. (Reprinted from Hempel (1965), pp. 245–290).
- Illari, P. (2013). Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, 78, 237–255.
- Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, 3, 438–444.
- Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78, 601–627.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: The University of Chicago Press.
- Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Moore, R. Y. (1997). Circadian rhythms: Basic neurobiology and clinical applications. *Annual Review of Medicine*, 48, 253–266.
- Moore-Ede, M. C. (1986). Physiology of the circadian timing system: Predictive versus reactive homeostasis. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, 250, 737–752.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- Salmon, W. (1989). Four decades of scientific explanation. In P. Kitcher & W. Salmon (Eds.), *Minnesota studies in the philosophy of science 13, scientific explanation* (pp. 3–219). Minneapolis: University of Minnesota Press.
- Stepp, N., & Turvey, M. T. (2010). On strong anticipation. *Cognitive Systems Research*, 11, 14–164.
- Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, 3, 425–437.
- Strevens, M. (2013). No understanding without explanation. *Studies in History and Philosophy of Science Part A*, 44(special section), 510–515.
- Thelen, E. (1995). Time-scale dynamics and the development of an embodied cognition. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 69–100). Cambridge, MA: MIT Press.
- Thelen, E., Schöner, G., Scheier, C., & Smith, S. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, 24, 1–34.
- Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science*, 69, 212–233.
- Trout, J. D. (2007). The psychology of explanation. *Philosophy Compass*, 2, 564–596.
- van Gelder, T. (1998). The dynamical hypothesis in cognitive science (with peer commentary). *Brain and Behavioral Sciences*, 21, 615–665.
- van Gelder, T., & Port, R. (1995). It's about time: An overview of the dynamical approach to cognition. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 1–43). Cambridge MA: MIT Press.
- van Leeuwen, M. (2005). Questions for the dynamicist. *Minds and Machines*, 15, 271–333.

- Voss, H. U. (2000). Anticipating chaotic synchronization. *Physical Review E*, 61, 5115–5519.
- Walmsley, J. (2008). Explanation in dynamical cognitive science. *Minds and Machines*, 18, 331–348.
- Waskan, J. (2011). Intelligibility and the cape: Combatting anti-psychologism about explanation. In *Epistemology of modeling & simulation: Building research bridges between the philosophical and modeling communities*. Pittsburgh. Retrieved April 8, 2014 from <http://philsci-archive.pitt.edu/view/confandvol/2011eomsbrbbpmcph13april2011.html>.
- Wheeler, M. (2005). *Reconstructing the cognitive world*. Cambridge, MA: MIT Press.
- Wimsatt, W. (1976). Reductive explanation: A functional account. In R. S. Cohen (Ed.), *PSA 1974* (pp. 671–710). Dordrecht: Reidel. (Reprinted from *Conceptual issues in evolutionary biology: an anthology* (1st ed.), pp. 477–508, by E. Sober, Ed., 1984, Cambridge MA: MIT Press.
- Woolley, D. W., & Shaw, W. (1954). A biochemical and pharmacological suggestion about certain mental disorders. *Proceedings of the National Academy of Sciences United States of America*, 40, 228–231.
- Wright, C. D. (2012). Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science*, 2, 375–394.
- Zednik, C. (2011). The nature of dynamical explanation. *Philosophy of Science*, 78, 238–263.